BELLCOMM, INC.
955 L'ENFANT PLAZA NORTH, S.W. WASHINGTON, D. C. 20024

B70 12063

SUBJECT: Preliminary Statistical Analysis of J-2 Engine Data - Case ~~340~~ 320

DATE: December 23, 1970

FROM: P. Gunther

ABSTRACT

The following are principal conclusions from a statistical analysis of J-2 engine dynamic gain tests performed at Huntsville and Rocketdyne facilities.

1. The Huntsville tests show poor repeatability.

2. The Rocketdyne data represents a different statistical population than the Huntsville data.

3. The scatter in the data for the two facilities is about the same.

4. It is not statistically valid to state that the scatter at low frequencies is less than at high frequencies.

RECEIVED
FEB 1971
NASA STI FACILITY
INPUT BRANCH

(NASA-CR-116222) PRELIMINARY STATISTICAL ANALYSIS OF J-2 ENGINE DATA (Bellcomm, Inc.) 23 p

N79-72579

Unclas
00/20 12815

FF No. 602

CR-116222 (NASA CR OR TMX OR AD NUMBER)

28 (CATEGORY)

[redacted]

BELLCOMM, INC.
955 L'ENFANT PLAZA NORTH, S.W. WASHINGTON, D. C. 20024

B70 12063

SUBJECT: Preliminary Statistical Analysis of J-2 Engine Data - Case 340

DATE: December 23, 1970

FROM: P. Gunther

MEMORANDUM FOR FILE

This note presents a statistical analysis of J-2 engine data from dynamic gain tests performed at Huntsville (H) and Rocketdyne (R) test facilities, as reported in the Rocketdyne letter report 70RC10500*. The investigation is in response to queries concerning (a) the compatability of the H and R data, and (b) whether the scatter at low frequencies is significantly smaller than at high frequencies (11 Hz is of interest for POGO). The gain data represents the ratio of chamber pressure to suction pressure resulting from pulsing the inlet pressure. Phase shift data is also available but will not be considered here**.

Data

Figures 1-10, reproduced from the Rocketdyne letter, plot gain vs. frequency for different mixture ratios and NPSH values. Individual regressions for Huntsville and Rocketdyne have been sketched in freehand. Qualitatively, the shapes of the regressions are extremely diverse, and the magnitudes of the R and H gains quite different. (See last two columns of Table 1.) More specifically,

1. Overall levels of gain differ significantly for R and H. In seven runs R has consistently larger values than H; in run 7, R is less; while in runs 5 and 9, with markedly different H and R regressions, there is some overlapping.

2. The R regression tends to be constant or increasing with frequency--only for runs 5 and 7 does a definite decrease appear. All H runs decrease with frequency (run 8 is U-shaped).

---

*"J-2 Engine Transfer Functions", by E. W. Larson, 30 September 1970.

**Cursory examination indicates that the phase data is considerably better behaved than the gain data.

Summary Calculations (Tables 1 and 2)

The frequency range was divided into five interval groupings: 1) 10-15 Hz, 2) 16-20, 3) 21-25, 4) 26-30, 5) 31-35. The data within each interval was combined and the mean and standard deviation (s.d.) calculated, for R and H separately as well as combined. For the four H runs with more than one replication, both the pooled and the unpooled results are shown in Table 1. Note that the R data does not contain frequencies greater than 25 Hz, and that none of the R tests were replicated.

A preliminary examination of the extreme s.d.'s prompted three adjustments to the R data. In run 1 a point showing positive db. gain was omitted--this decreased the s.d. from 1.82 to 1.37. In runs 5 and 7, where abrupt changes in gain occurred, an end point of one interval was transferred to the adjacent interval. The s.d.'s decreased from 1.76 to 1.18 and from 2.03 to .46, respectively.

For the H data the effect was investigated of transferring isolated points (the only one in its class interval) to the adjacent interval. Of 14 such transfers 10 led to some improvement. Three transfers resulted in marked increases in s.d. Since the latter affected the total scatter disproportionately, it was decided not to employ transfers in the H analysis.

The analysis of variance for all 10 tests combined (after correcting for individual regressions) is summarized in Table 2, which contains most of the numerical quantities referred to below. Details of this analysis are explained in the Appendix.

Replications of Huntsville Runs (Figure A)

Before considering the compatability of the H and R data, it is natural first to inquire about the repeatability of the H tests alone. There were two runs with three replications each and two runs with two replications each. Figure A plots the s.d.'s, both individually and pooled, for each of the five frequency groupings. Pooling of runs 1 and 6, each with two replications, does not give significantly different results. This shows up in Figure A with the pooled s.d.'s intermediate to the unpooled values. In run 3 one of the replications had much larger gain than the remaining two (see Figure 3 and Table 1). In Figure A, the pooled s.d., which includes the variation between means, is seen to exceed the unpooled s.d.'s. In run 9 this effect occurs only at frequencies above 25 Hz; moreover, at low frequency (10-15 Hz) the data comes mainly from one replication, but the extremely large scatter--the s.d. of 2 was the largest of any run--turns out to dominate the overall numerical results for H.

In summary, the reproducibility of the H data is not good enough. At least one, and possibly two, of the four replicated runs is definitely from a different "population". This fact tends to cast suspicion on the adequacy of the unreplicated runs.

<u>Variation in Scatter for R and H Separately</u> (Figure B)

Figure B plots the s.d.'s vs. frequency (interval) for each of the 10 runs. The R data at the lowest frequency has an average s.d. (weighted according to degrees of freedom) of .66 whereas the two high frequency intervals have averages exceeding unity. Thus the R data shows significantly smaller scatter at lower frequencies.

The s.d.'s for the H data are far less consistent with no discernable trend. The overall unpooled s.d. for the H data has about the same magnitude as for R, approximately .9. Pooling the replicated tests increases the average H s.d. to 1.3. The low frequency s.d. has the greatest value (1.33 unpooled), due primarily to the erratic behavior of test number 9 noted above. There is also an unusually low value (.52) for the 25-30 Hz interval.

It is believed that much of the scatter in the H data may be contributed by the sharply decreasing regression. A refined procedure would estimate scatter assuming a linear regression for each interval.

The main point is that the H and R data exhibit opposite effects regarding the scatter at low vs. high frequencies.

<u>H and R Data Combined</u> (Figure C)

Figure C plots the combined scatter s.d. for each frequency interval (up to 25 Hz) when the H and R deviations are taken about the common mean. The s.d. at low frequency is seen to be less than at the higher frequencies. This appears to be somewhat at odds with the conclusion of the preceding paragraph, and can be explained either by the fact that only 4 of the 10 H runs contained 10-15 Hz data, or by the fact that for these runs, there was smaller divergence between the H and R regressions.

Conclusions

Statistical tests of significance bearing on each of the following conclusions are contained in the Appendix.

1. The repeatability of the H tests is unsatisfactory, especially test number 3. The variation between means of replicated runs is almost three times what would be expected from identical processes.

2. The R test data represents a different population from the H data. The regressions differ in shape as well as magnitude, and the trend of the scatter, as frequency increases, goes in opposite directions.

3. The overall scatter for R (.9) is approximately the same as for H when replicates are not pooled.

4. For the combined H and R data, the smaller scatter at low frequencies (1.3) in comparison with higher frequencies (1.5 and 1.9), is more apparent than real. In part it reflects the fact that Huntsville conducted relatively few tests at low frequencies. In addition, the variance of the combined data are artificially inflated at the higher frequencies, because the divergence between H and R regressions is greatest there.

P Gunther

1031-PG-cp P. Gunther

APPENDIX

ANALYSIS OF VARIANCE

The analysis of variance is a formal statistical procedure for decomposing the total sum of squares (SS). Let i denote facility, H or R; j the replication (rep for short) for H ($j=1,\ldots,r$); and k the variate ($k=1,\ldots,n_{ij}$). To facilitate interpretation, each $\sum$ expression below is followed by a descriptive phrase with SS additionally subscripted by the appropriate row number in Table 2. The usual subscript dot notation for mean is employed to indicate the subscripts over which the average is taken. The basic identity is

$$\sum(x_{ijk}-x_{...})^2 = \sum(x_{ijk}-x_{ij.})^2 + \sum(x_{ij.}-x_{i..})^2 + \sum(x_{i..}-x_{...})^2$$

SS_6(Total R+H variation) = SS_7(within:H unpooled and R) + SS_2(between H rep means) + SS_5(between H and R means)

Since R has only a single replicate, we have

$$SS_2 = \sum(X_{ij.}-x_{i..})^2 = \sum_{j=1}^{r} n_{Hj}(x_{Hj.}-x_{H..})^2$$

$$\sum(x_{ijk}-x_{ij.})^2 = \sum_{k=1}^{n_R}(x_{Rk}-x_{R.})^2 + \sum_{j=1}^{r}\sum_{k=1}^{n_{Hj}}(x_{Hjk}-x_{Hj.})^2$$

SS_7(within:H unpooled +R) = SS_4(R) + SS_1(within H reps: unpooled)

In addition, we have the decomposition

$$\left\{ \begin{aligned} &\sum(x_{Hjk}-x_{H..})^2 = \sum_{k,j} (x_{Hjk}-x_{Hj.})^2 + \sum_{j} n_{Hj}(x_{Hj.}-x_{H..})^2 \\ &\qquad SS_3(\text{H pooled}) = SS_1(\text{H unpooled}) + SS_2(\text{between H rep. means}) \end{aligned} \right.$$

whence

$$SS_6 = SS_1 + SS_2 + SS_4 + SS_5 = SS_3 + SS_4 + SS_5$$

Also of interest is

$$\sum(x_{ij.}-x_{i..}) = SS_8(\text{within:H pooled+R}) = SS_3(\text{H pooled}) + SS_4(\text{R})$$

Note finally that SS_5(between H and R) is given by

$$\sum(x_{i..}-x_{...})^2 = n_R(x_{R.}-x_{...})^2 + n_{H.}(x_{H..}-x_{...})^2 = \frac{n_R n_{H.}}{n_R+n_{H.}}(x_{R.}-x_{H..})^2.$$

Using the values in Table 1, the above analysis of variance can be readily carried out, for each test condition and each frequency interval. (When there are no multiple H replications, the analysis simplifies.) The results for the total of the 10 cases are shown in Table 2.

The values in Table 2 can be used to perform F-ratio tests of significance for comparing two (independent) estimates of variance, assuming that the data are normally distributed. If v_1 and v_2 represent the degrees of freedom in the numerator and denominator, respectively, F is determined by

$$F_{v_1, v_2} = \frac{SS_1}{v_1} \Big/ \frac{SS_2}{v_2}$$

$$= \frac{s_1^2}{s_2^2}$$

Presented below are several such tests that relate to the four main conclusions of the analysis. The probabilities corresponding to the F-values were roughly estimated from the limited tabulated values of the F distribution in "Biometrika Tables for Statisticians". A small value of P (associated with a large value of F) denotes a significant difference in the two estimates of variance, i.e., the difference is unlikely to be the result merely of random sampling. When F=1, P=50% (but = 100% for 2-sided test).

1* <u>Significance of H reps</u>

$$F_{25,145} = \frac{\text{H reps}}{H_{UP}} = \left(\frac{2.61}{.89}\right)^2 = 8.55$$

P<.1% -- significant

*The verbal description of F identifies the appropriate SS (or, more accurately, the mean square) listed in the first column of Table 2. P denotes pooled and UP unpooled. The numerical values of S used in 1-3 are those shown in the total column.

2. Significance of H vs. R means

$$\begin{cases} F_{24,225} = \dfrac{\text{Between H and R}}{H_P + R} = \left(\dfrac{3.73}{1.22}\right)^2 = 9.29 \\ P<.1\% \text{ -- significant} \end{cases}$$

Using H_{UP} in place of H_P leads to even greater significance. To test whether H rep differences in 1. are comparable to H vs R:

$$\begin{cases} F_{24,25} = \dfrac{\text{Between H and R}}{\text{H reps}} = \left(\dfrac{3.73}{2.61}\right)^2 = 2.04 \\ P \simeq 4.2\% \text{ -- marginally significant} \end{cases}$$

3. Comparison of H vs. R scatter

a. H unpooled

$$\begin{cases} F_{107,145} = \dfrac{R}{H_{UP}} = \left(\dfrac{.915}{.89}\right)^2 = 1.05 \\ P \simeq 80\% \text{ (2-sided) -- not significant} \end{cases}$$

b. H pooled

$$\begin{cases} F_{170,107} = \dfrac{H_P}{R} = \left(\dfrac{1.30}{.915}\right)^2 = 2.0 \\ P<.1\% \text{ (1-sided) -- significant} \end{cases}$$

4. Low vs. high frequency scatter

a. Comparison of H vs. R at low frequency

$$\left\{\begin{array}{l} F_{13,41} = \dfrac{(H_{UP})_{10*}}{(R)_{10}} = \left(\dfrac{1.33}{.66}\right)^2 = 4.02 \\ \\ P<.2\% \text{ (2-sided) -- significant} \end{array}\right.$$

b. Comparison of H + R

$$\left\{\begin{array}{l} F_{66,54} = \dfrac{(H_{UP}+R)_{16}}{(H_{UP}+R)_{10}} = \left(\dfrac{.98}{.87}\right)^2 = 1.73 \\ \\ P\approx 2.5\% \text{ -- fairly significant} \end{array}\right.$$

$$\left\{\begin{array}{l} F^{**}_{72,59} = \dfrac{(H_{P}+R)_{10}}{(H_{P}+R)_{16}} = \left(\dfrac{1.24}{1.17}\right)^2 = 1.135 \\ \\ P\approx 30\% \text{ -- not significant} \end{array}\right.$$

$$\left\{\begin{array}{l} F_{82,63} = \dfrac{(H+R)_{16}}{(H+R)_{10}} = \left(\dfrac{1.49}{1.31}\right)^2 = 1.295 \\ \\ P\approx 12\% \text{ -- not significant} \end{array}\right.$$

---

*Subscript denotes lower limit of frequency interval(s).

**Note that the s.d. for 10-15 Hz is greater than for 16-20 Hz.

$$\begin{cases} F_{186,63} = \dfrac{(H+R)_{16+21}}{(H+R)_{10}} = \left|\dfrac{(561.67/186)}{(1.31)^2}\right| = 1.77 \\ \\ P \simeq .5\% \text{ -- significant} \end{cases}$$

To test whether H vs. R mean differences at low frequency are comparable to high frequencies:

$$F_{20,4} = \frac{(\text{Between H and R})_{16+21}}{(\text{Between H and R})_{15}} = \left|\frac{(317/20)}{2.0^2}\right| = 3.96$$

$P \simeq 10\%$ -- not significant

Table 1. SUMMARY OF DATA CHARACTERISTICS

				10 - 15 Hz			16 - 20			21 - 25			26 - 30			31 - 36				
Figure	MR	NPSH	Run	n	$\overline{x}$ *	s	n	$\overline{x}$	s	n	$\overline{x}$	s	n	$\overline{x}$	s	n	$\overline{x}$	s	Regression Shape	R-H Magn.
1	4.5	45	Ha	4	4.4	1.22	5	6.3	.63	5	5.7	.56	2	4.1	1.24	3	6.2	1.78		
			Hb				5	5.4	.82	5	5.8	.99	2	5.4	.53	2	7.	.21		
			HT	4	4.4	1.22	10	5.8	.83	10	5.7	.76	4	4.8	1.06	5	6.5	1.33	Non-linear	
			R	6	3.8	.35	5	2.9	.85	4[1]	3.0	1.37							Linear incr'g	+
			R+H	10	4.0	.80	15	4.8	1.65	14	4.9	1.53								
2	4.5	50	H				1	8.9	0	4	9.0	.66	5	11.7	.77	1	.7	0	Linear decr'g	
			R	5	5.0	.78	5	4.0	1.04	5	4.6	.90							Constant	+
			R+H				6	4.9	2.19	9	6.5	2.58								
3	4.5	55	Ha	1	9.4	0	5	8.9	.53	5	9.4	1.31	2	11.3	.48	3	12.6	1.65	Decr'g	
			Hb	5	4.2	1.20	5	5.7	.92	5	4.5	.87							Constant	
			Hc	4	7.9	.40	3	8.4	1.16	3	9.4	.52	2	10.1	.60				Decr'g	
			HT	10	6.2	2.29	13	7.6	1.74	13	7.5	2.64	4	10.7	.83	3	12.6	1.65		+
			R	6	5.6	.96	5	5.2	1.66	5	5.3	1.15							Constant	
			R+H	16	6.0	1.88	18	6.9	1.98	18	6.9	2.51								
4	4.5	60	H				1	9.3	0	4	10.7	.47	5	12.2	.49	1	12.6	0	Linear decr'g	
			R	6	7.1	.52	5	7.1	.92	5	6.3	.45							Constant	+
			R+H				6	7.5	1.20	9	8.3	2.36								

Table 1 (con'd)

Figure	MR	NPSH	Run	10 - 15 Hz n	$\bar{x}$	s	16 - 20 n	$\bar{x}$	s	21 - 25 n	$\bar{x}$	s	26 - 30 n	$\bar{x}$	s	31 - 36 n	$\bar{x}$	s	Regression Slope	R-H Magn.
5	5.0	50	H				1	6.7	0	4	7.2	1.39	4	7.5	.47	2	9.7	.57	Linear decr'g	
			R	5^2	6.9	1.18	6^2	3.9	.66	5	6.3	1.19							Non-linear	+?
			R+H				7	4.3	1.21	9	6.7	1.29								
6	5.0	55	Ha	1	5.6	0	5	5.9	.80	5	7.8	.33	2	8.3	.14	3	10.7	1.04		
			Hb	1	4.7	0	5	5.9	1.26	5	7.4	.51	2	8.4	.02	3	10.7	1.25		
			HT	2	5.2	.64	10	5.9	1.00	10	7.6	.44	4	8.3	.10	6	10.7	1.03	Linear decr'g	
			R	6	6.0	.27	4	5.3	1.63	5	4.4	.76							Incr'g	+
			R+H	8	5.8	.53	14	5.7	1.18	15	6.5	1.64								
7	5.0	60	H				1	7.5	0	4	9.1	.79	4	10.2	.12	2	12.3	.42	Linear decr'g	
			R	6	6.3	.44	3^2	7.2	.46	5^2	13.2	1.37							Non-linear decr'g	-
			R+H				4	7.3	.40	9	11.4	2.40								
8	5.5	50	H				1	7.1	0	4	8.9	.90	4	9.2	.32	2	8.2	.11	Non-linear U	
			R	4	8.0	.64	3	6.8	.32	3	6.2	.49							Linear incr'g	+
			R+H				4	6.9	.28	7	7.7	1.62								
9	5.5	55	Ha	1	7.1	0	4	7.4	.23	5	8.1	.55	2	8.8	.11	3	10.8	.75		
			Hb	1	8.5	0	5	8.6	1.30	4	8.7	.87	2	8.0	.25	2	8.5	.50		
			Hc	4	5.3	2.03	3	8.3	.76	3	8.4	.40	1	8.2	0					
			HT	6	6.1	2.07	12	8.2	1.02	12	8.3	.65	5	8.3	.43	5	9.9	1.37	Decr'g	
			R	4	8.5	.46	2	9.5	.78	3	6.6	1.23							Non-linear U	+
			R+H	10	7.1	1.98	14	8.3	1.07	15	8.0	1.03								

Table 1 (con'd)

Figure	MR	NPSH	Run	10 - 15 Hz			16 - 20			21 - 25			26 - 30			31 - 36			Regression Slope	R-H Magn.
				n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s		
10	5.5	60	H				1	9.6	0	4	10.0	.98	4	10.3	.47	2	11.3	1.34	Linear decr'g	
			R	3	8.7	.40	3	8.7	1.15	5	7.4	.69							Increasing	+
			R+H				4	9.0	1.03	9	8.6	1.55								

*For simplicity, the minus sign is omitted from the $\bar{x}$ values (db)

[1]Extreme point deleted.

[2]Border point transferred to adjacent interval.

TABLE 2.[+] ANALYSIS OF VARIANCE (TOTAL FOR 10 TESTS)

	10-15 Hz			16-20			21-25			26-30			31-35			Total		
	SS	df	s	SS	df	s	SS	df	s	SS	df	s	SS	df	s	SS	df	s
(1) H unpooled	23.07	13	1.33	28.45	35	.91	35.60	53	.82	7.48	28	.52	20.76	16	1.14	115.36	145	.89
(2) H reps	50.60	5	3.18	34.33	6	2.39	74.62	6	3.53	3.72	5	.86	6.90	3	1.52	170.17	25	2.61
(3) H pooled: (1)+(2)	73.67	18	2.02	62.78	41	1.24	110.22	59	1.37	11.20	33	.58	27.66	19	1.21	285.53	170	1.30
(4) R	17.95	41	.66	35.68	31	1.07	35.91	35	1.01							89.54	107	.915
(5) Between H and R	16.05	4	2.00	83.08	10	2.88	234.00	10	4.84							333.13	24	3.73
(6) H+R: (3)+(4)+(5)	107.67	63	1.31	181.54	82	1.49	380.13	104	1.91							669.34*	249	1.64
(7) H unpooled+R: (1)+(4)	41.02	54	.87	64.13	66	.98	71.51	88	.90							176.66*	208	.92
(8) H pooled+R: (3)+(4)	91.62	59	1.24	98.46	72	1.17	146.13	94	1.25							336.21*	225	1.22

*H data only from 10-25 Hz included

[+]See Appendix for notation and definitions.

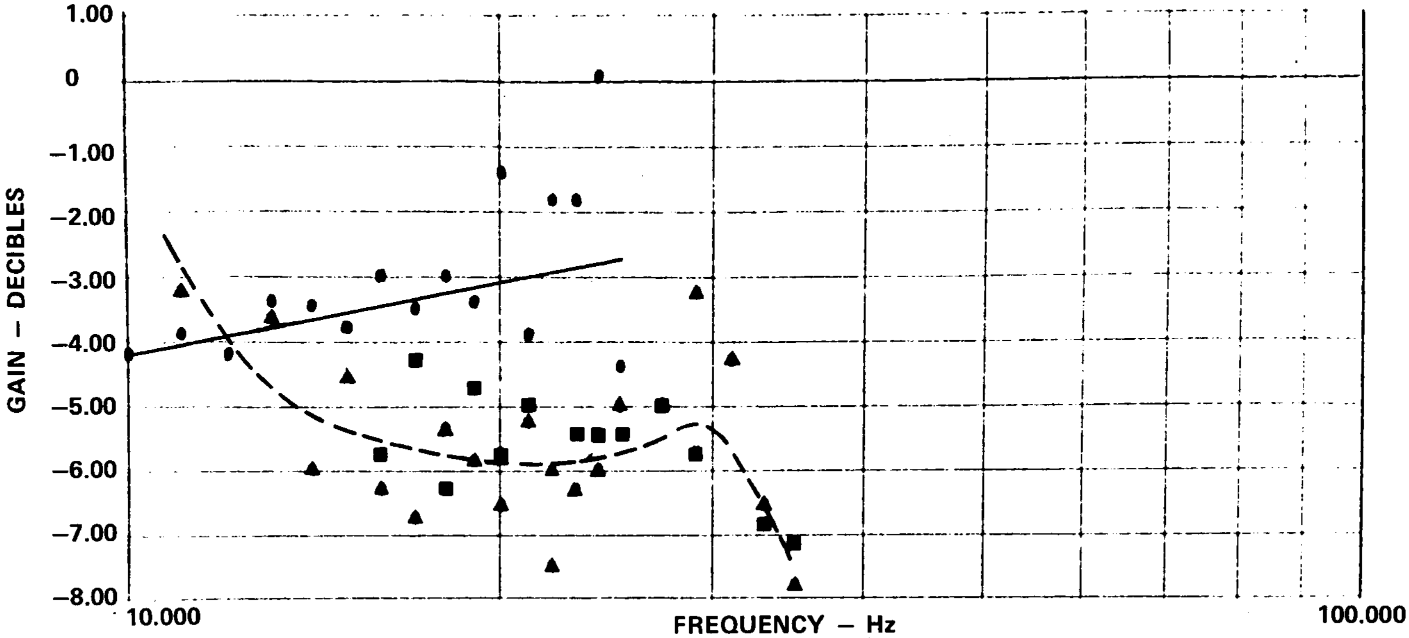

FIGURE 1B - MIXTURE RATIO 4.5, NPSH 45

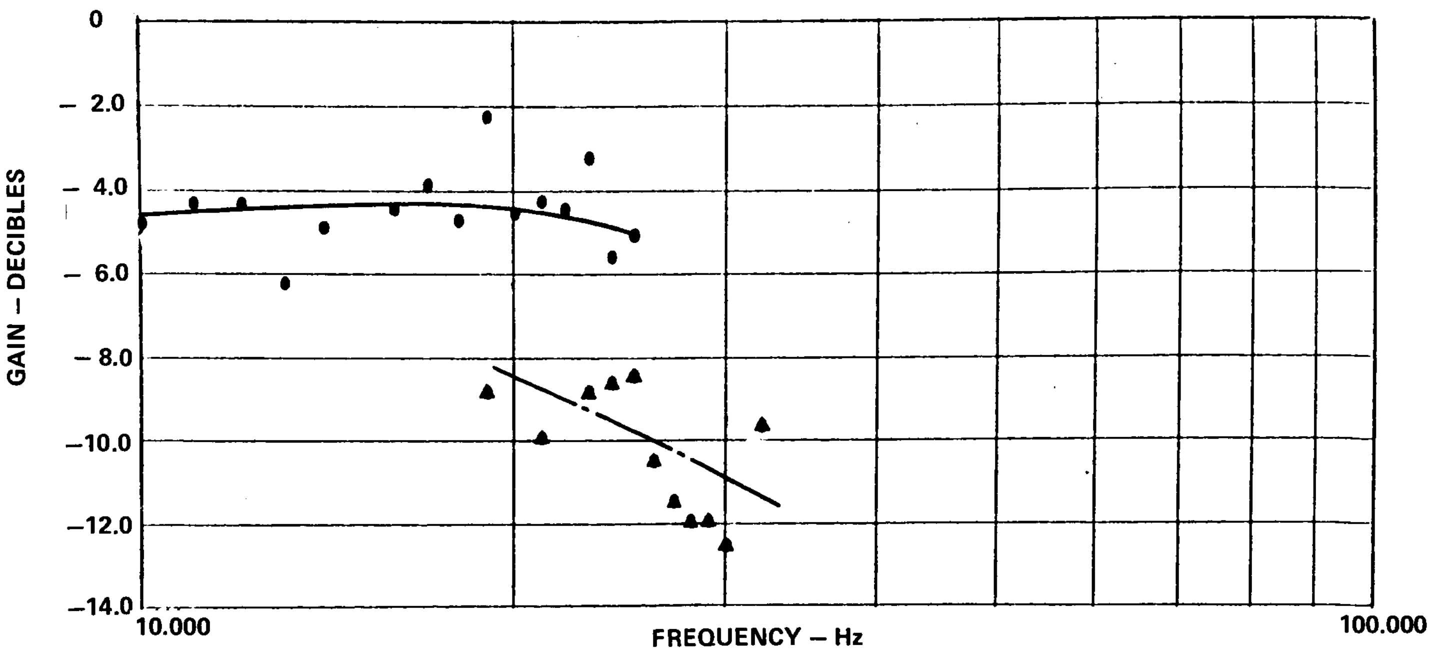

FIGURE 2B - MIXTURE RATIO 4.5, NPSH 50

LEGEND:

HUNTSVILLE ▲ ■

ROCKETDYNE ●

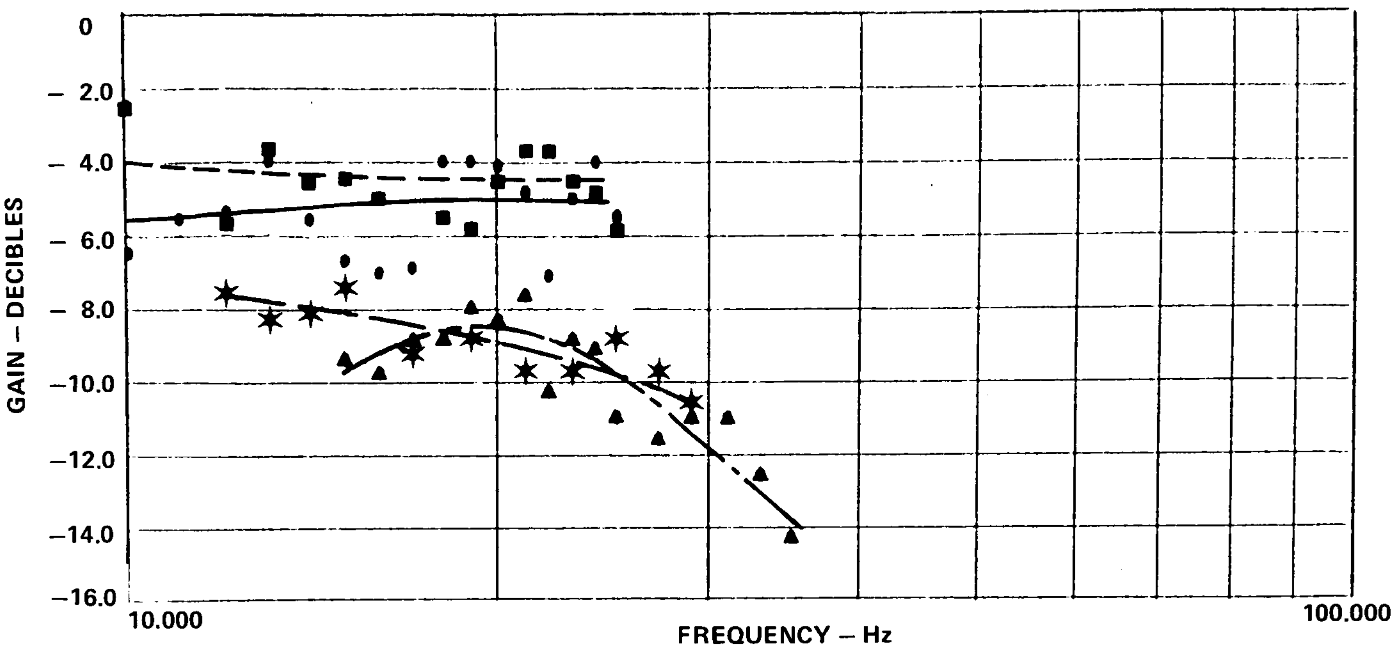

FIGURE 3B - MIXTURE RATIO 4.5, NPSH 55

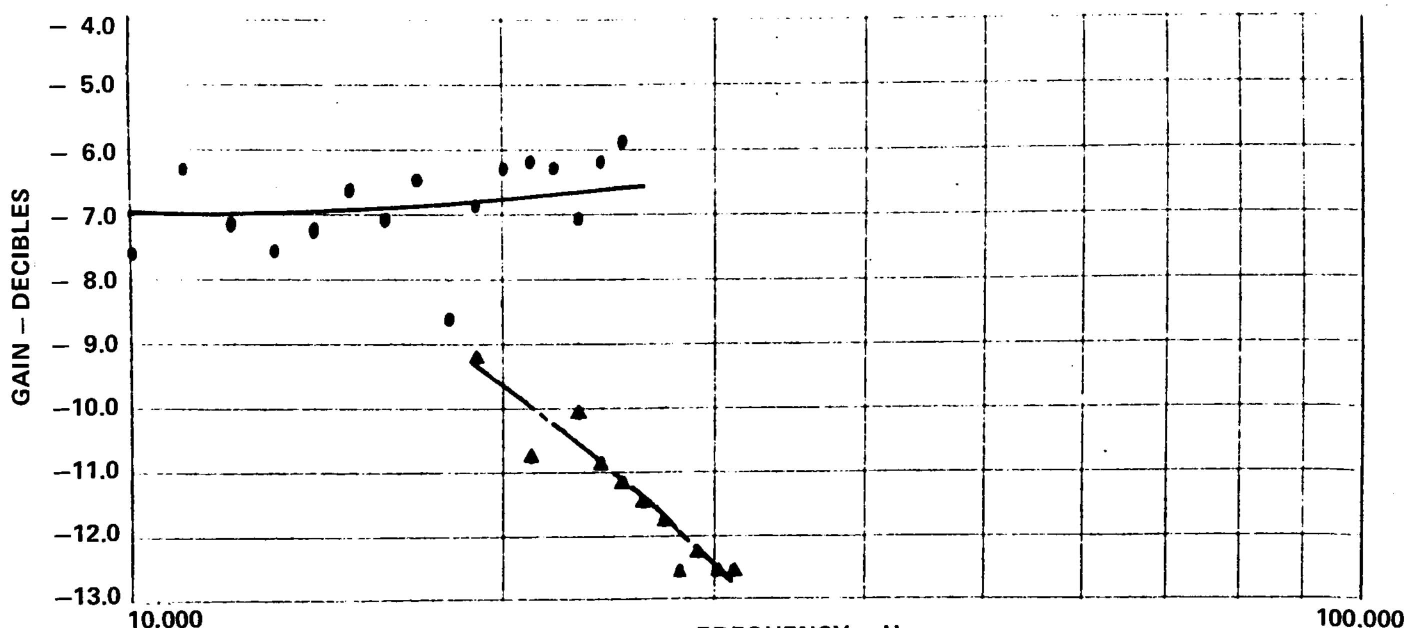

FIGURE 4B - MIXTURE RATIO 4.5, NPSH 60

LEGEND:

HUNTSVILLE ▲ ■ ✶

ROCKETDYNE ●

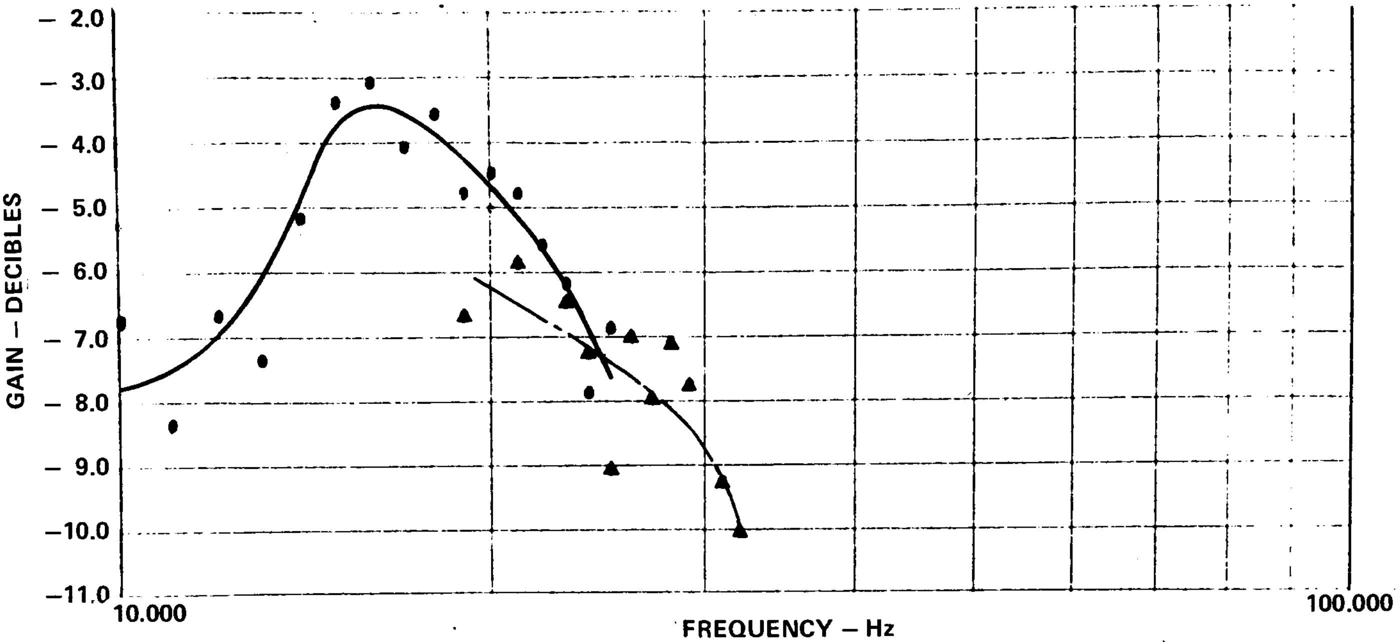

FIGURE 5B - MIXTURE RATIO 5.0, NPSH 50

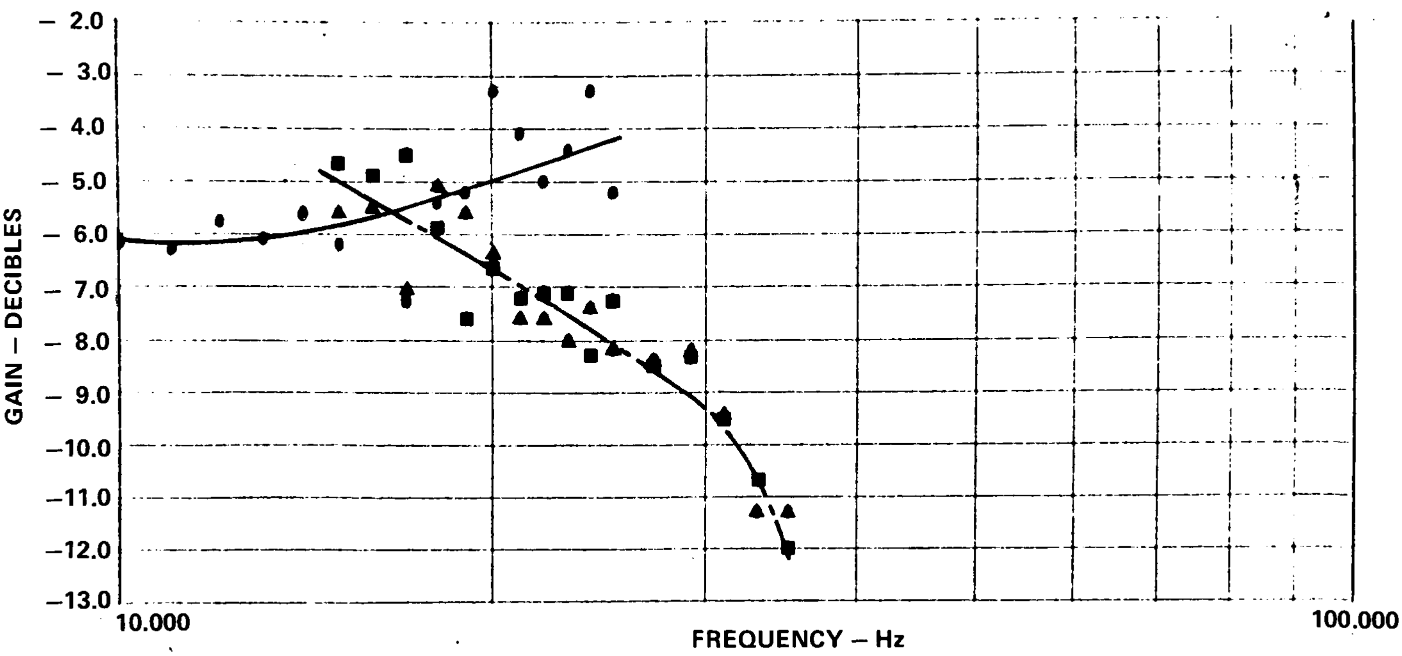

FIGURE 6B - MIXTURE RATIO 5.0, NPSH 55

LEGEND:

HUNTSVILLE ▲ ■

ROCKETDYNE ●

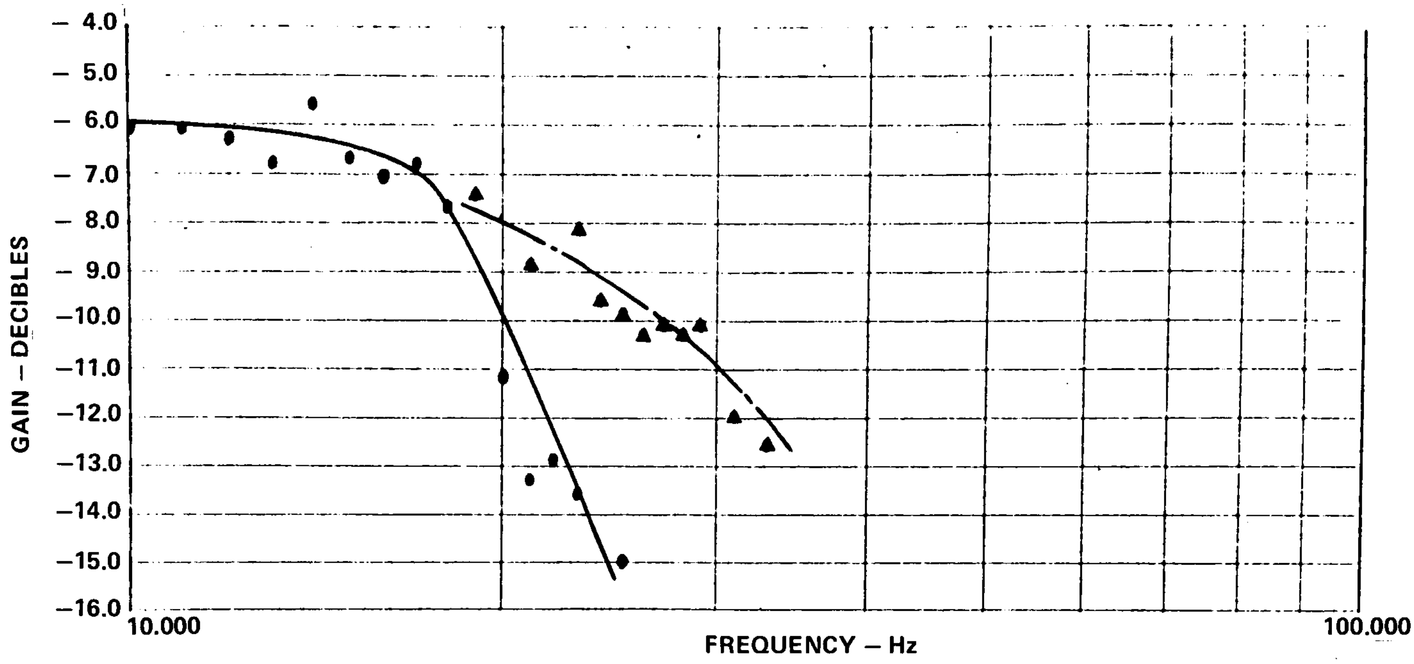

FIGURE 7B - MIXTURE RATIO - 5.0, NPSH 60

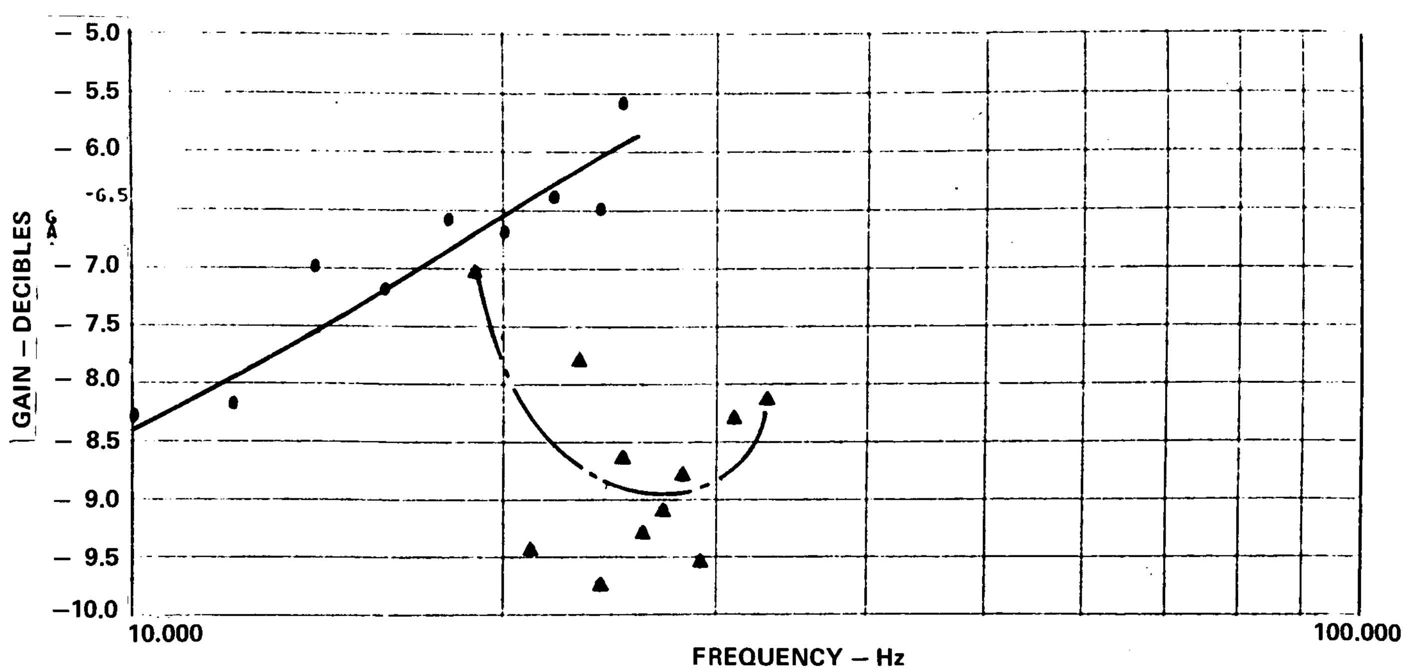

FIGURE 8B - MIXTURE RATIO - 5.5, NPSH 50

LEGEND:

HUNTSVILLE ▲■

ROCKETDYNE ●

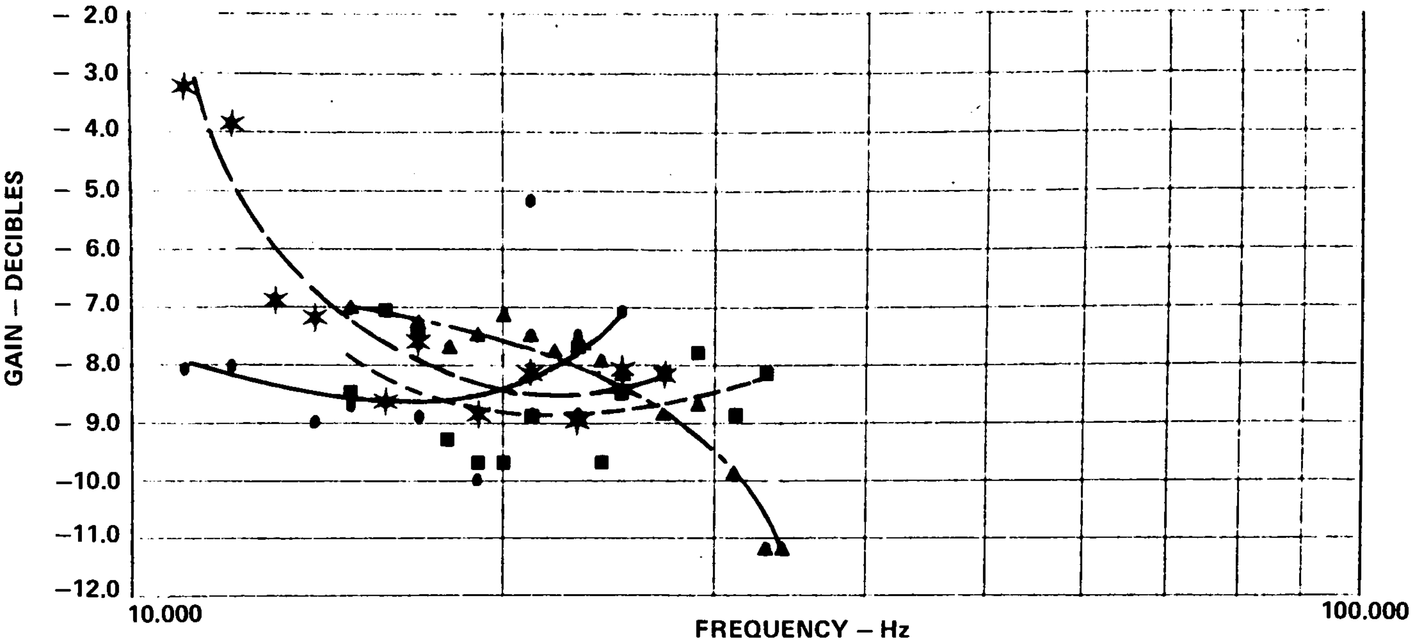

FIGURE 9B - MIXTURE RATIO 5.5, NPSH 55

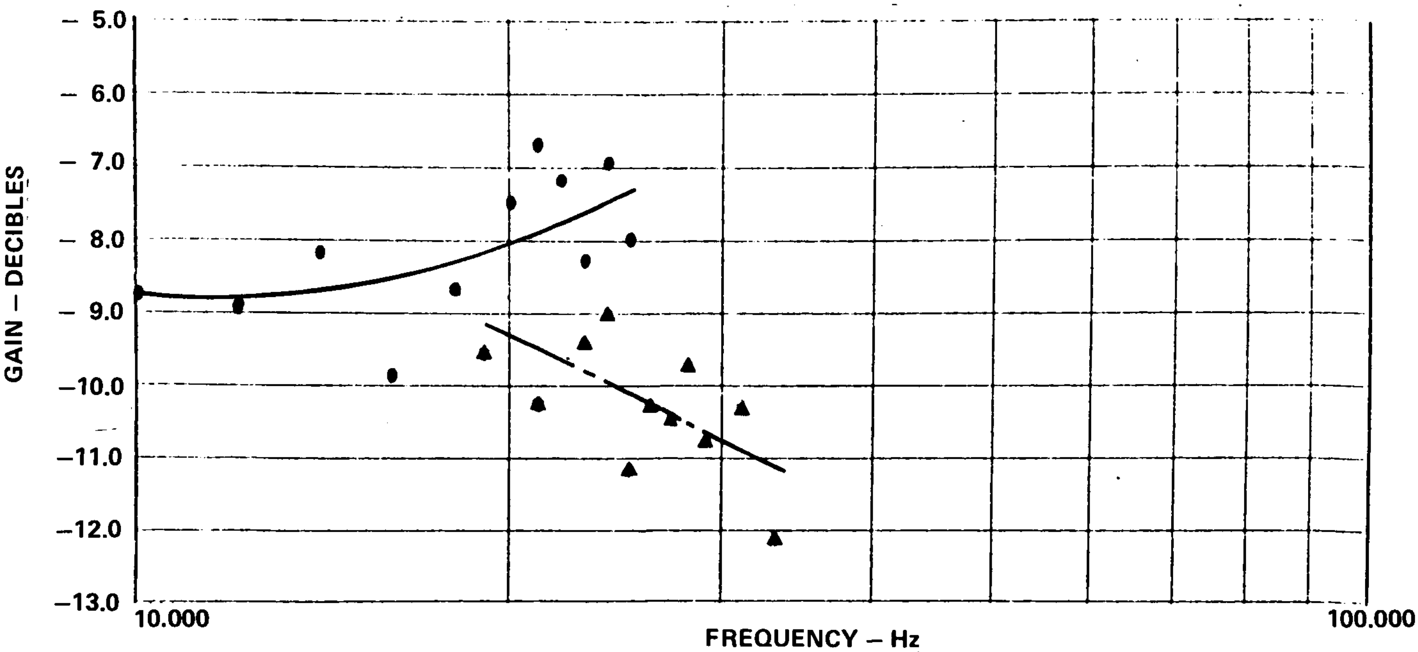

FIGURE 10B - MIXTURE RATIO 5.5, NPSH 60

LEGEND:
HUNTSVILLE ▲ ■ ✶
ROCKETDYNE ●

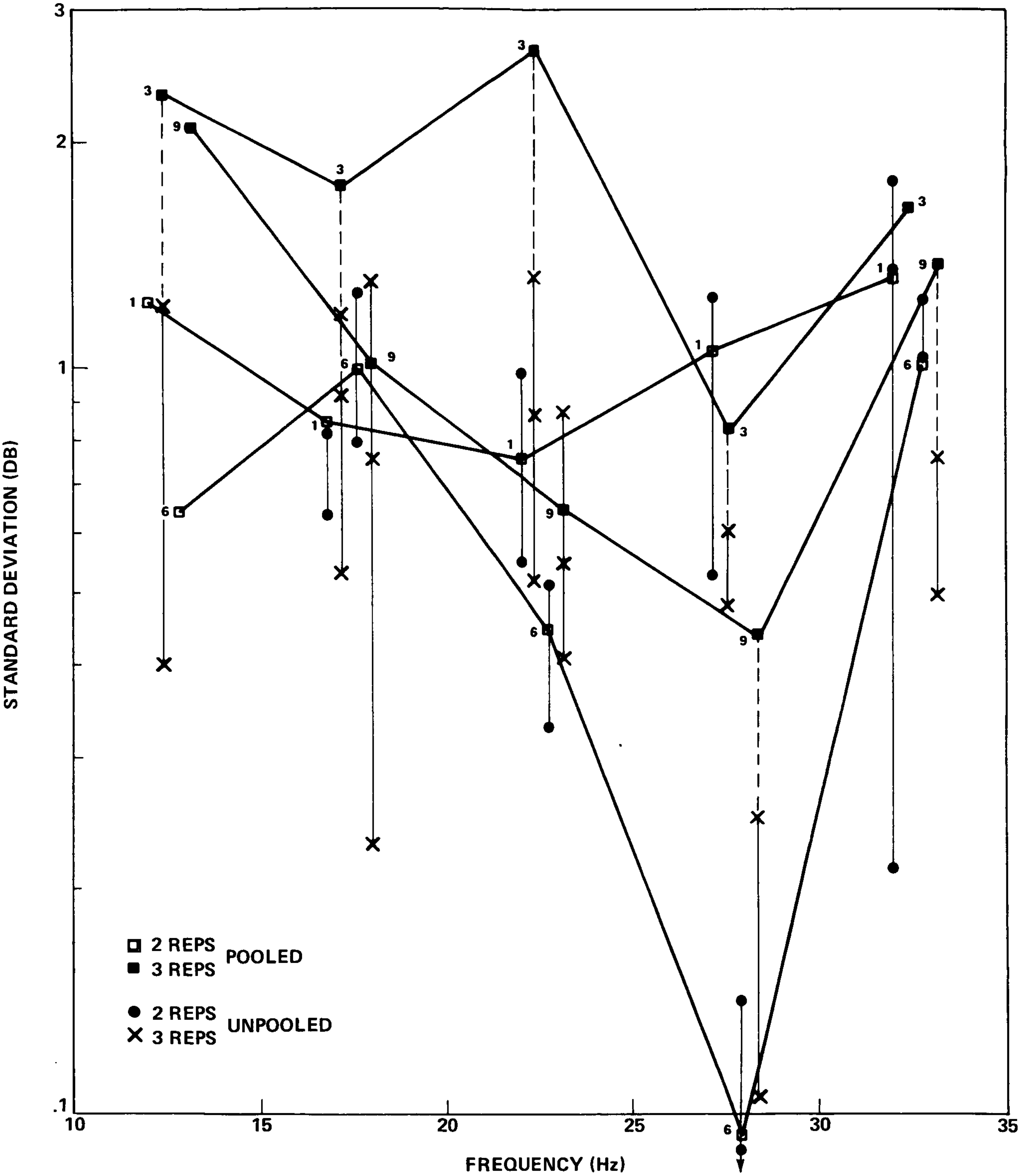

FIGURE A - REPLICATED HUNTSVILLE TESTS — POOLED VS. UNPOOLED SCATTER

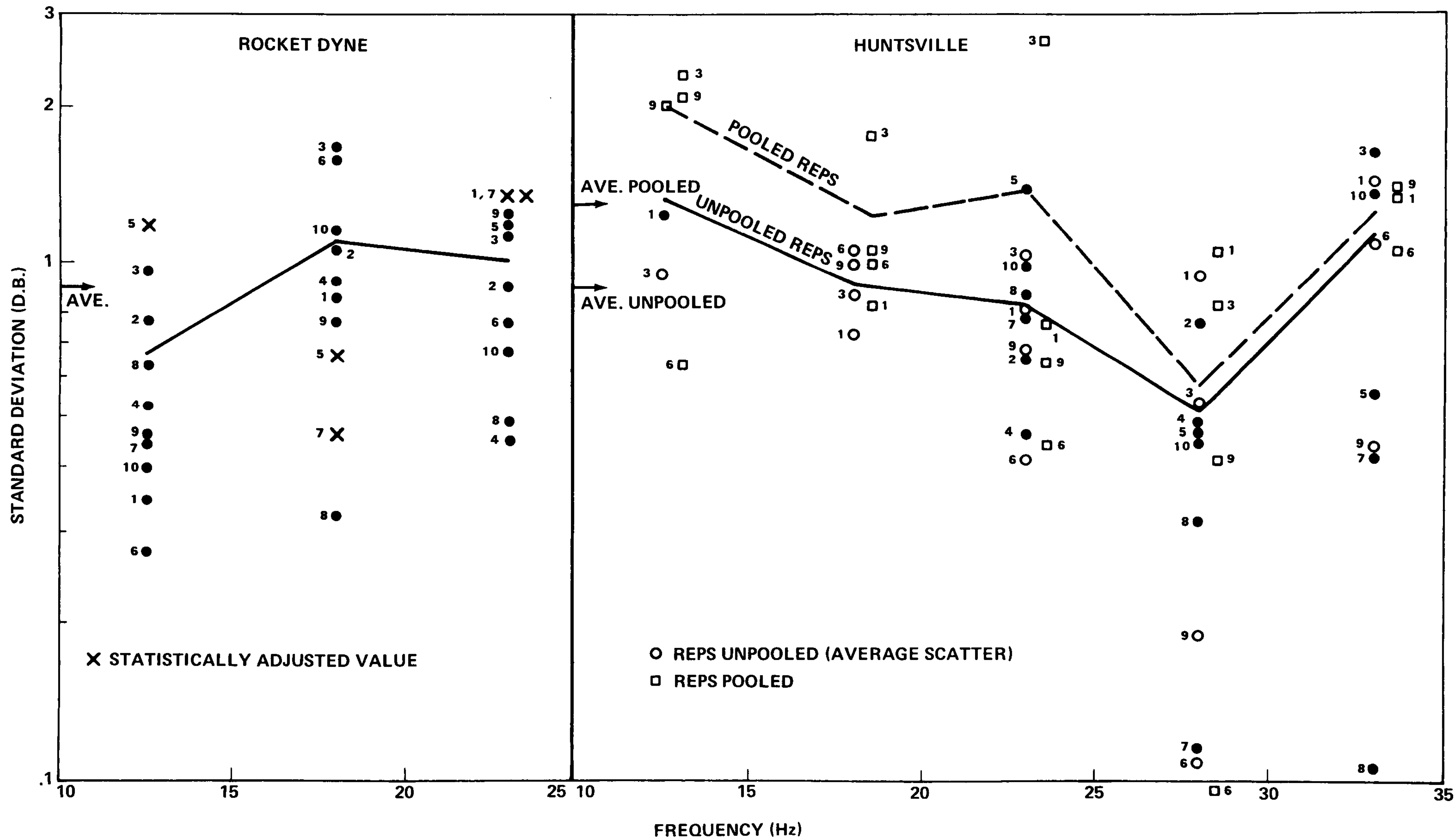

FIGURE B. VARIATION OF SCATTER WITH FREQUENCY

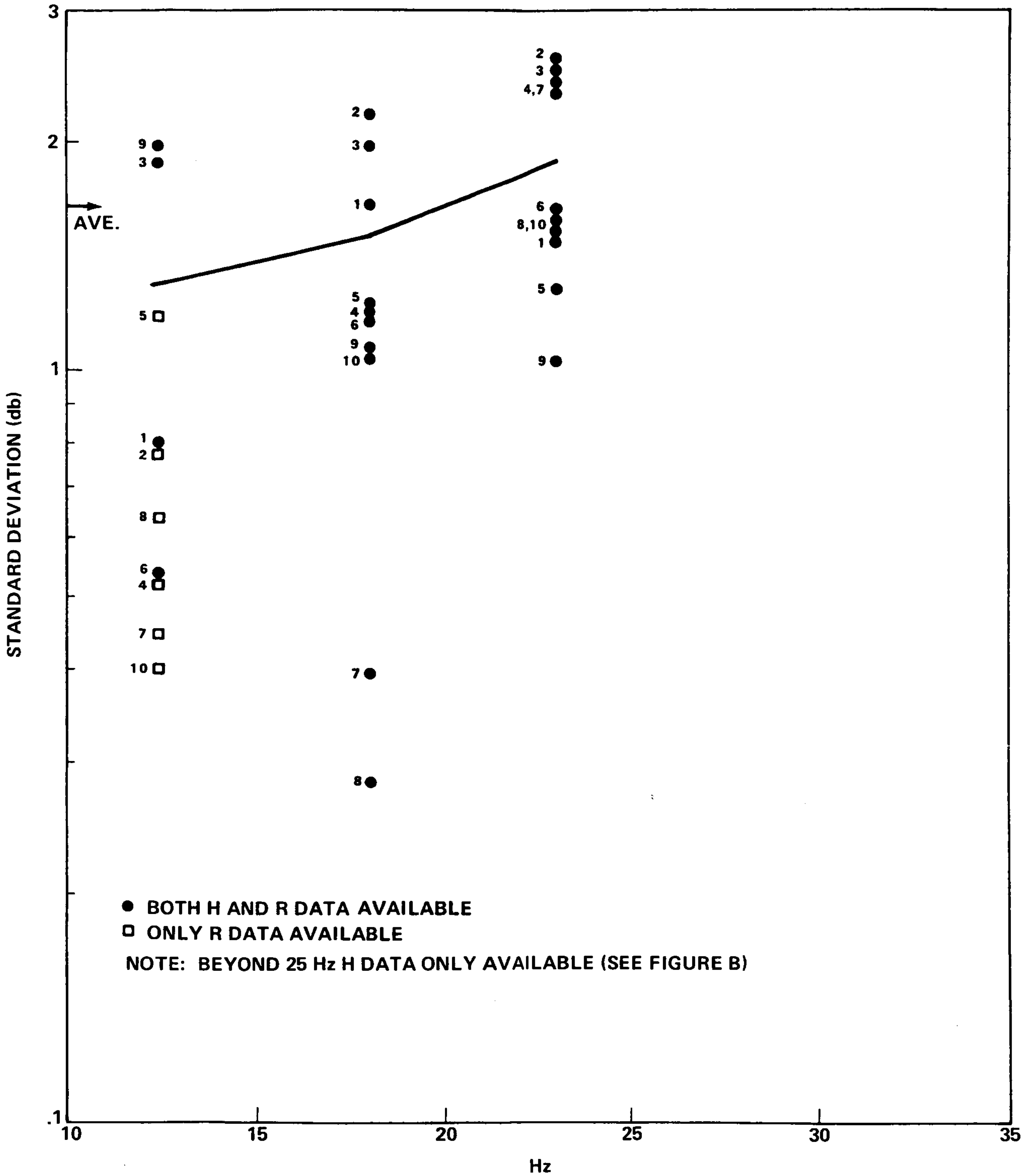

FIGURE C - VARIATION OF SCATTER WITH FREQUENCY — HUNTSVILLE AND ROCKETDYNE DATA COMBINED